Energy Optimization across Training and Data for Multiuser Minimum Sum-MSE Linear Precoding

Adam J. Tenenbaum and Raviraj S. Adve
Dept. of Electrical and Computer Engineering, University of Toronto
10 King's College Road, Toronto, Ontario, M5S 3G4, Canada
Email: {adam,rsadve}@comm.utoronto.ca

***Abstract*—This paper considers minimum sum mean-squared error (sum-MSE) linear transceiver designs in multiuser downlink systems with imperfect channel state information. Specifically, we derive the optimal energy allocations for training and data phases for such a system. Under MMSE estimation of uncorrelated Rayleigh block fading channels with equal average powers, we prove the separability of the energy allocation and transceiver design optimization problems. A closed-form optimum energy allocation is derived and applied to existing transceiver designs. Analysis and simulation results demonstrate the improvements that can be realized with the proposed design.**

I. Introduction

Transceiver designs that minimize the sum of mean squared errors (sum-MSE) under a sum power constraint in the multiuser downlink with full channel state information (CSI) at the base station are well researched [**?**], [**?**], [**?**], [**?**]. In these papers, an uplink-downlink duality is used to transform a nonconvex downlink problem into an equivalent convex virtual uplink problem. Recent studies [**?**], [**?**], [**?**] have extended these original papers to the case of imperfect CSI, deriving an MSE duality in the presence of channel estimation errors and providing robust transceiver designs.

In order to design precoders, the base station must obtain estimates of the channel coefficients. If channel reciprocity holds (i.e. the uplink and downlink channels are statistically identical), these estimates can be provided by training in the uplink (e.g., using uplink sounding, as in the WiMAX standard [**?**]). However, in frequency division duplex systems (and in some broadband time division duplex systems [**?**]), channel reciprocity does not apply. In this case, channel estimation must be performed in the downlink and communicated back to the base station using an uplink feedback mechanism. In this paper, we consider imperfect CSI estimation at the mobile receivers, but assume that the imperfect estimates are also available at the base station (via an error-free and delay-free feedback mechanism)[1].

The algorithms designed in [**?**], [**?**], [**?**] for minimization of the sum-MSE under a sum-power constraint presume that fixed channel estimation error variances σ_k^2 are provided by a predetermined estimation mechanism. In this paper, we address the problem of jointly designing a training sequence for MMSE CSI estimation and designing linear transceivers for minimum sum-MSE communication. We consider the optimum allocation of limited available energy between the training and data communication phases for each quasi-static communication block.

In Section II, we describe the channel model under consideration and review the design of training sequences for MMSE channel estimation. We then present the linear precoding system model and provide an overview of the design of minimum sum-MSE linear precoders with imperfect CSI and fixed transmit power. In Section III, we formulate the joint design problem for energy allocation and precoder design. We present a closed-form solution for the optimum training energy, and apply the result to existing precoder design techniques. Performance and behaviour of the proposed approach are illustrated in Section IV, and we draw conclusions in Section V. Appendix A derives the MMSE channel estimation error variance and the calculations of our main proof are presented in Appendix B.

Notation: We use the following conventions: italics represent scalars, lower case boldface type is used for vectors, and upper case boldface represents matrices, (e.g., $x, \mathbf{x}, \mathbf{X}$, respectively). Entries in vectors and matrices are denoted as $[\mathbf{x}]_i$ and $[\mathbf{X}]_{i,j}$. The superscripts ^{T} and ^{H} denote the transpose and Hermitian operators. $\mathbb{E}[\cdot]$ represents the statistical expectation operator while $\mathbf{I}_N$ is the $N \times N$ identity matrix. $\|\mathbf{x}\|_1$ and $\|\mathbf{x}\|_2$ denote the 1-norm (sum of entries) and Euclidean norm. $\mathrm{diag}(\mathbf{x})$ represents the diagonal matrix formed using the entries in vector $\mathbf{x}$, and $\mathrm{diag}\left[\mathbf{X}_1, \ldots, \mathbf{X}_k\right]$ is the block diagonal concatenation of matrices $\mathbf{X}_1, \ldots, \mathbf{X}_k$. The $\mathrm{vec}(\mathbf{X})$ operator stacks the columns of the matrix $\mathbf{X}$ in a single vector. $\mathcal{CN}(\mathbf{m}, \mathbf{R})$ denotes the complex multivariate Gaussian probability distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{R}$.

II. System Model and Background

A. Channel Model

In the linear precoding system illustrated in Fig. 1, a base station with M antennas transmits to K decentralized mobile users with N_k antennas each over flat wireless channels. The

[1] In this regard, this work complements [**?**], where we consider perfect receiver CSI estimates and a feedback mechanism incorporating prediction, error, and delay.

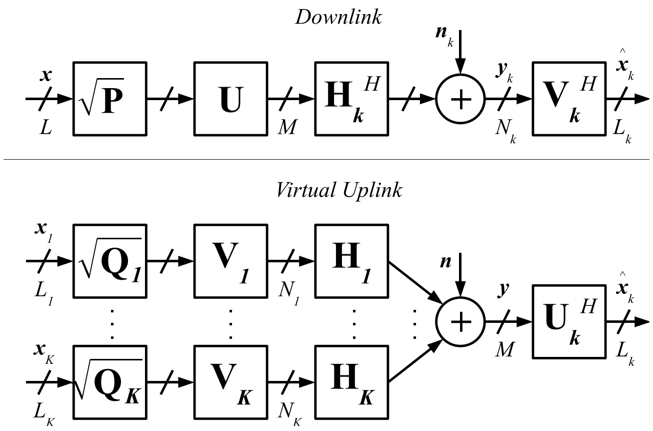

Fig. 1. Data processing for user k in downlink and virtual uplink.

channel between the transmitter and user k is represented by the $N_k \times M$ matrix $\mathbf{H}_k^H$, and the overall $N \times M$ channel matrix is $\mathbf{H}^H$, with $\mathbf{H} = [\mathbf{H}_1, \ldots, \mathbf{H}_K]$, and where $N = \sum_k N_k$ is the total number of receive antennas in the system. We assume that all channel coefficients are i.i.d. and drawn from a zero-mean complex Gaussian distribution with variance σ_H^2; that is, $\text{vec}(\mathbf{H}) \sim \mathcal{CN}(\mathbf{0}, \sigma_H^2 \mathbf{I}_{MN})$. We consider a quasi-static (block fading) channel model, where the channel coefficients are assumed to be fixed for a coherence interval of n consecutive symbol periods. The first n_T transmissions in each block are training symbols which the mobile receivers use to estimate the downlink channel, $\hat{\mathbf{H}}_k^H$; these imperfect CSI estimates are assumed to be available at the base station via an error-free and delay-free feedback mechanism. We consider the stochastic error model (as used in [**?**], [**?**], [**?**]) where the true channel is modelled as a sum of the estimated channel and an independent additive error term, $\mathbf{H}_k = \hat{\mathbf{H}}_k + \mathbf{E}_k$ with $\text{vec}(\mathbf{E}_k) \sim \mathcal{CN}(\mathbf{0}, \sigma_k^2 \mathbf{I}_{MN_k})$, and $\mathbf{E} = [\mathbf{E}_1, \ldots, \mathbf{E}_K]$.

B. MMSE Channel Estimation and Training

Training sequence and estimator design can be simplified under the assumption of uncorrelated channel coefficients by considering training for vector channels from the M transmit antennas to each individual receive antenna. To simplify notation in this section, we consider training for a single vector channel $\mathbf{h}^H$. Channel estimation is performed by transmitting a set of n_T training signal vectors, $\mathbf{X}_T = [\mathbf{x}_{T,1}, \ldots, \mathbf{x}_{T,n_T}]$, from the M transmit antennas without precoding. $n_T \geq M$ training symbol vectors must be sent to guarantee resolvability of the individual channel coefficients. The received signal vector is $\mathbf{y}_T = \mathbf{h}^H \mathbf{X}_T + \mathbf{z}$, where $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_{n_T})$, and the MMSE channel estimate $\hat{\mathbf{h}}_{\text{MMSE}}^H = \mathbf{y}_T \mathbf{A}_0$ is found using the linear MMSE estimator $\mathbf{A}_0 = \left(\mathbf{X}_T^H \mathbf{X}_T + \frac{\sigma_n^2}{\sigma_H^2} \mathbf{I}_{n_T} \right)^{-1} \mathbf{X}_T^H$. Under the sum energy constraint, $\text{tr}\left[\mathbf{X}_T^H \mathbf{X}_T\right] \leq E_T$, where E_T is the energy allocated to training, and the assumption of independent channel coefficients, a sufficient condition for optimality of the training matrix is $\mathbf{X}_T \mathbf{X}_T^H = \frac{E_T}{M} \mathbf{I}_M$ [**?**]; that is, we are free to select any training matrix with orthogonal rows. When using the MMSE estimator, there is no benefit using any more than $n_T = M$ training symbols. For algorithmic simplicity, we choose the set of training vectors $\mathbf{X}_T = \sqrt{\frac{E_T}{M}} \mathbf{I}_M$. One may also choose $\mathbf{X}_T$ as the scaled size-M DFT matrix, $[\mathbf{X}_T]_{m,n} = \frac{\sqrt{E_T}}{M} e^{-j2\pi mn/M}$, which has the additional benefit of balancing training power equally over each transmit antenna in each training symbol.

In Appendix A, we show that the estimation errors of each channel coefficient are equal under the assumption of i.i.d. channels with variance σ_H^2, taking the value

$$\sigma_e^2 = \left(\sigma_H^{-2} + \frac{1}{\sigma_n^2} \frac{E_T}{M} \right)^{-1}. \quad (1)$$

As we illustrate in Section II-D, the assumption of equal estimation error variance is critical in maintaining convexity of the virtual uplink sum-MSE minimization problem.

C. Linearly Precoded Data Communication Model

Following training, we assume that all of the remaining $n_D = n - M$ symbol periods in each block will be used to broadcast data symbols. Under the block fading assumption, the channel $\mathbf{H}$ does not change during these n_D data transmissions; thus, we can design a single precoder/decoder pair to be used for all transmissions in the block. It follows that the remaining available energy to be used for data $(E_D = E_{\max} - E_T)$ should be divided equally over the n_D data transmissions, resulting in a maximum per-symbol transmit power $P_D = (E_{\max} - E_T)/n_D$.

During each data transmission, user k receives L_k data symbols $\mathbf{x}_k = [x_{k1}, \ldots, x_{kL_k}]^T$ from the base station, and the vector $\mathbf{x} = \left[\mathbf{x}_1^T, \ldots, \mathbf{x}_K^T\right]^T$ comprises independent symbols with unit average energy ($\mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right] = \mathbf{I}_L$, where $L = \sum_{k=1}^K L_k$). User k's data streams are precoded by the $M \times L_k$ transmit filter $\mathbf{U}_k = [\mathbf{u}_{k1}, \ldots, \mathbf{u}_{kL_k}]$, where $\mathbf{u}_{kj}$ is the precoding beamformer for stream j of user k with $\|\mathbf{u}_{kj}\|_2 = 1$, and the precoders are combined in the $M \times L$ global transmitter precoder matrix $\mathbf{U} = [\mathbf{U}_1, \ldots, \mathbf{U}_K]$. Power is allocated to user k's data streams in the vector $\mathbf{p}_k = [p_{k1}, \ldots, p_{kL_k}]^T$ and $\mathbf{P}_k = \text{diag}\left[\mathbf{p}_k\right]$; we define the downlink power allocation matrix as $\mathbf{P} = \text{diag}\left[\mathbf{p}_1^T, \ldots, \mathbf{p}_K^T\right]$ with $\text{tr}\left[\mathbf{P}\right] \leq P_D$. Based on this model, user k receives a length-N_k vector $\mathbf{y}_k^{DL} = \mathbf{H}_k^H \mathbf{U} \sqrt{\mathbf{P}} \mathbf{x} + \mathbf{n}_k$, where the superscript ^{DL} indicates the downlink, and $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_{N_k})$. To estimate its L_k symbols $\mathbf{x}_k$, user k applies the $L_k \times N_k$ receive filter $\mathbf{V}_k^H$, yielding the estimated symbols $\hat{\mathbf{x}}_k^{DL} = \mathbf{V}_k^H \mathbf{H}_k^H \mathbf{U} \sqrt{\mathbf{P}} \mathbf{x} + \mathbf{V}_k^H \mathbf{n}_k$.

In order to design the sum-MSE minimizing precoder for the downlink, we use the virtual uplink, also illustrated in Fig. 1, where each matrix is replaced by its conjugate transpose. We emphasize that the virtual uplink is only a mathematical construct to be used for precoder design, and that its use does not require reciprocity of the true uplink and downlink channels. We imagine that transmissions from mobile user k in the virtual uplink propagate via the *flipped channel* $\mathbf{H}_k$ to the base station. The transmit and receive filters for user k become $\mathbf{V}_k$ and $\mathbf{U}_k^H$ respectively, with normalized precoding beamformers; i.e., $\|\mathbf{v}_{kj}\|_2 = 1$, and the uplink precoder matrices are gathered as a block diagonal matrix $\mathbf{V} = \mathrm{diag}\left[\mathbf{V}_1, \ldots, \mathbf{V}_K\right]$. Power is allocated to user k's data streams as $\mathbf{q}_k = \left[q_{k1}, \ldots, q_{kL_k}\right]^T$, with $\mathbf{Q}_k = \mathrm{diag}\left[\mathbf{q}_k\right]$, $\mathbf{Q} = \mathrm{diag}\left[\mathbf{q}_1^T, \ldots, \mathbf{q}_K^T\right]$, and $\mathrm{tr}\left[\mathbf{Q}\right] \leq P_D$. The received symbol vector at the base station and the estimated symbol vector for user k are $\mathbf{y}^{UL} = \mathbf{H}\mathbf{V}\sqrt{\mathbf{Q}}\mathbf{x}+\mathbf{n} = \sum_{i=1}^K \mathbf{H}_i\mathbf{V}_i\sqrt{\mathbf{Q}_i}\mathbf{x}_i+\mathbf{n}$ and $\hat{\mathbf{x}}_k^{UL} = \mathbf{U}_k^H\mathbf{y}^{UL}$, respectively, with $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2\mathbf{I}_M)$.

D. Robust Convex Minimum Sum-MSE Precoder Design

The MSE matrix for user k in the virtual uplink can be written as

$$\begin{aligned}\varepsilon_k^{UL} &= \mathbb{E}_{\mathbf{E},\mathbf{x},\mathbf{n}}\left[\left(\hat{\mathbf{x}}_k^{UL} - \mathbf{x}_k\right)\left(\hat{\mathbf{x}}_k^{UL} - \mathbf{x}_k\right)^H\right] \\ &= \mathbb{E}_{\mathbf{E}}\Big[\mathbf{U}_k^H\left(\mathbf{H}\mathbf{V}\mathbf{Q}\mathbf{V}^H\mathbf{H}^H + \sigma_n^2\mathbf{I}\right)\mathbf{U}_k \\ &\qquad -\mathbf{U}_k^H\mathbf{H}_k\mathbf{V}_k\sqrt{\mathbf{Q}_k} - \sqrt{\mathbf{Q}_k}\mathbf{V}_k^H\mathbf{H}_k^H\mathbf{U}_k + \mathbf{I}_{L_k}\Big] \\ &= \mathbf{U}_k^H\tilde{\mathbf{R}}\mathbf{U}_k - \mathbf{U}_k^H\hat{\mathbf{H}}_k\bar{\mathbf{V}}_k - \bar{\mathbf{V}}_k^H\hat{\mathbf{H}}_k^H\mathbf{U}_k + \mathbf{I}_{L_k},\end{aligned} \tag{2}$$

where $\bar{\mathbf{V}}_k = \mathbf{V}_k\sqrt{\mathbf{Q}_k}$, $\tilde{\mathbf{R}} = \hat{\mathbf{H}}\mathbf{V}\mathbf{Q}\mathbf{V}^H\hat{\mathbf{H}}^H + \sigma_{\text{eff}}^2\mathbf{I}_M$. Here, we have defined the effective noise power $\sigma_{\text{eff}}^2 = \sigma_n^2 + \sum_{k=1}^K \sigma_k^2 \mathrm{tr}\left[\mathbf{V}_k\mathbf{Q}_k\mathbf{V}_k^H\right]$, under the general model with different estimation error variances σ_k^2 for each user k. We have also assumed the independence of data symbols, noise, and estimation errors. The optimum robust virtual uplink receiver for user k is found using the MMSE (Wiener) filter $\tilde{\mathbf{U}}_k^H = \bar{\mathbf{V}}_k^H\hat{\mathbf{H}}_k^H\tilde{\mathbf{R}}^{-1}$. The resulting (minimum) sum-MSE is

$$\begin{aligned}\mathrm{SMSE}_{UL} &= \sum_{k=1}^K L_k - \mathrm{tr}\left[\tilde{\mathbf{R}}^{-1}\sum_{k=1}^K \hat{\mathbf{H}}_k\bar{\mathbf{V}}_k\bar{\mathbf{V}}_k^H\hat{\mathbf{H}}_k^H\right] \\ &= L - M + \sigma_{\text{eff}}^2\mathrm{tr}\left[\tilde{\mathbf{R}}^{-1}\right]\end{aligned} \tag{3}$$

which follows from $\mathrm{tr}\left[\mathbf{AB}\right] = \mathrm{tr}\left[\mathbf{BA}\right]$, linearity of the trace operator, and the definition of $\tilde{\mathbf{R}}$. Since the beamforming vectors $\mathbf{v}_{kj}$ have unit norm, it follows that $\mathrm{tr}\left[\mathbf{V}_j\mathbf{Q}_j\mathbf{V}_j^H\right] = \sum_{l=1}^{L_j} q_{jl} = \|\mathbf{q}_j\|_1$ is the sum of powers allocated to user j's data streams. Under a sum-power constraint with a maximum transmit power of P_D, the non-convex virtual uplink sum-MSE minimization problem can be formally defined as

$$\begin{aligned}(\mathbf{V}^*, \mathbf{Q}^*) = &\underset{\mathbf{V},\mathbf{Q}}{\arg\min}\left(\sigma_n^2 + \sum_{k=1}^K \sigma_k^2\|\mathbf{q}_k\|_1\right)\mathrm{tr}\left[\tilde{\mathbf{R}}^{-1}\right] \\ \text{s.t.}\quad & q_{kl} \geq 0 \quad k = 1, \ldots, K;\ \ l = 1, \ldots, L_k, \\ & \mathrm{tr}\left[\mathbf{Q}\right] \leq P_D.\end{aligned} \tag{4}$$

When the channel estimation error variances are equal ($\sigma_k^2 = \sigma_e^2$), the effective noise becomes $\sigma_{\text{eff}}^2 = \sigma_n^2 + \sigma_e^2\sum_k \|\mathbf{q}_k\|_1$. Since the minimum sum-MSE is a non-increasing function of $\sum_k \|\mathbf{q}_k\|_1$, we can assume that all available power allocated to data transmission will be used [**?**]. Thus, the effective noise can be further simplified as $\sigma_{\text{eff}}^2 = \sigma_n^2 + \sigma_e^2 P_D$ for the optimum precoder, which is no longer a function of the uplink power allocations q_{kl}. The optimization problem (4) thus becomes convex (the minimization of $\mathrm{tr}\left[\tilde{\mathbf{R}}^{-1}\right]$ under a sum power constraint), and can thus be solved using the algorithm from [**?**] designed for the perfect CSI case by substituting the effective noise σ_{eff}^2 for the noise term σ_n^2 in the original design.

III. Joint Optimization of Energy and Precoder Design

The previous section describes the design of a robust minimum sum-MSE precoder for a fixed data power allocation, P_D. In this section, we extend this result by jointly optimizing the available training and data energy with the precoder design. As explained in Section II-C, the optimum strategy for sharing the available data energy E_D over n_D transmitted symbols is with equal energy in each transmission. Using this strategy, and substituting the estimation error variance from (1) into the effective noise variance, we define the joint optimization problem

$$\begin{aligned}(\mathbf{V}^*, \mathbf{Q}^*, E_T^*) = &\underset{\mathbf{V},\mathbf{Q},E_T}{\arg\min}\,\sigma_{\text{eff}}^2\mathrm{tr}\left[\tilde{\mathbf{R}}^{-1}\right] \\ \text{s.t.}\qquad & q_{kl} \geq 0 \quad k = 1, \ldots, K;\ \ l = 1, \ldots, L_k, \\ & \mathrm{tr}\left[\mathbf{Q}\right] = P_D, \quad P_D = \frac{E_{\max} - E_T}{n_D}, \\ & \sigma_{\text{eff}}^2 = \sigma_n^2 + \frac{P_D}{\left(\sigma_H^{-2} + \frac{1}{\sigma_n^2}\frac{E_T}{M}\right)}.\end{aligned} \tag{5}$$

Theorem 1: The optimum training energy E_T^* is

$$E_T^* = \begin{cases}\frac{E_{\max}\sqrt{M} - \frac{\sigma_n^2}{\sigma_H^2}M\sqrt{n_D}}{\sqrt{n_D}+\sqrt{M}} & E_{\max} > \frac{\sigma_n^2}{\sigma_H^2}\sqrt{Mn_D} \\ 0 & \text{otherwise.}\end{cases} \tag{6}$$

Proof: See Appendix B.

Corollary 1: The optimization of training/data energy allocation and the optimum precoder design in problem (5) are separable problems. This result can be seen directly in (6), as the optimum value of E_T is neither a function of $\mathbf{V}$ nor $\mathbf{Q}$.

Corollary 2: The sum-MSE minimizing precoder can be designed using existing algorithms by setting the sum power constraint $\mathrm{tr}\left[\mathbf{Q}\right] \leq P_D = \left(E_{\max} - E_T\right)/n_D$ and the noise power term to the effective noise power $\sigma_{\text{eff}}^2 = \sigma_n^2 + \sigma_e^2 P_D$.

Corollary 3: No information can be communicated using the proposed algorithm in the case where $E_{\max} \leq \frac{\sigma_n^2}{\sigma_H^2}\sqrt{Mn_D}$. If the total available energy fails to exceed this threshold, there is zero energy allocated to training; as a result, the estimated channel is $\hat{\mathbf{H}} = \mathbf{0}$ and the resulting symbol estimates are $\hat{\mathbf{x}}^{DL} = \mathbf{0}$ as well. It is difficult to provide an intuitive understanding of this result, as we do not have a closed-form expression for the minimum sum-MSE as a function of E_T; however, we have observed in simulations that when $E_{\max}$ falls below the threshold, the resulting minimum sum-MSE is an increasing function of E_T. It follows that the "best" (i.e., sum-MSE minimizing) strategy is to avoid training.

We can reinterpret this threshold result in the context of average received SNR. If we define the average transmitted

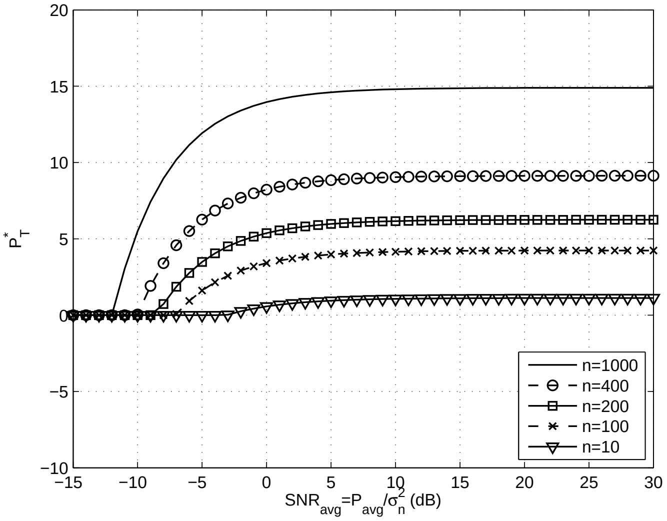

Fig. 2. Optimum training power P_T^* for varying block length n

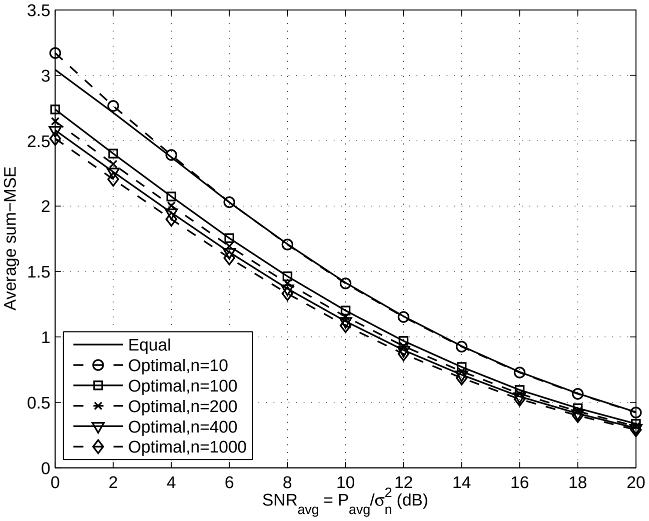

Fig. 3. Sum-MSE performance for equal and optimal energy allocations

power as $P_{\mathrm{avg}} \doteq E_{\max}/n$, we can rewrite the constraint as

$$SNR_{\mathrm{rx}} \doteq \frac{P_{\mathrm{avg}}\sigma_H^2}{\sigma_n^2} \le \frac{\sqrt{Mn_D}}{n_D + M}. \qquad (7)$$

It follows that as $n \to \infty$, a strictly positive optimum training power allocation is always feasible. Furthermore, the largest average received SNR value that the threshold can take on is $SNR_{\mathrm{rx}} = -3\mathrm{dB}$, corresponding to the maximum value of the RHS of (7) when $n_D = M$.

IV. Numerical Examples

We now present both analytical and simulation results to illustrate the behaviour and performance of the proposed algorithm. In these results, the flat Rayleigh fading channels are modelled with $\sigma_H^2 = 1$. We scale the total energy $E_{\max}$ proportionally to the block-length n to reflect a realistic average power constraint, $P_{\mathrm{avg}} = E_{\max}/n = \alpha$; in these simulations, we illustrate the case of $\alpha = 1$. As such, we define the average transmit SNR as $P_{\mathrm{avg}}/\sigma_n^2$, and find different SNR values by varying the noise power σ_n^2. These preliminary results illustrate performance in a system with $K = 2$ users, $M = 4$ base station antennas, and $N_1 = N_2 = L_1 = L_2 = 2$ receive antennas and data streams per user.

Figure 2 illustrates how the optimum power allocated to training, P_T^*, grows with average SNR and with block length n. We observe that as n grows, the optimum power allocated to training becomes significantly larger than the equal power allocation $P_T = 1$; however, P_T^* converges fairly rapidly with increasing SNR. We also observe the threshold behaviour described in Corollary 3.

Figures 3 and 4 illustrate the sum-MSE and average BER performance of the proposed algorithm. Results in each of these plots are generated using 5000 channel realizations per average SNR value, and data symbols are generated as uncoded QPSK. Here, we compare performance of the proposed algorithm to the case where equal power is allocated to both training and data symbols (i.e. $P_T = P_D = 1$). We observe notable performance improvements for large block lengths ($n \gg M$), with approximately 3 dB of SNR gain for $n = 1000$.

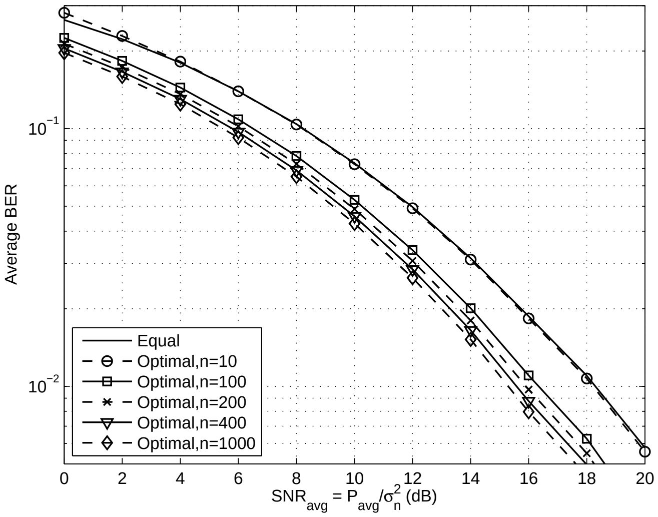

Fig. 4. Average BER performance for equal and optimal energy allocations

V. Conclusions

In this paper, we have considered the problem of allocating energy to training and data symbols for systems using minimum sum-MSE linear precoding in the multiuser MIMO downlink. We have derived the optimum closed-form energy allocation for the case of MMSE channel estimation when all users have statistically identical channels. Furthermore, we have proven separability of the energy allocation and precoder designs; thus, existing algorithms for minimum sum-MSE precoding can be applied following energy optimization. Preliminary simulation results demonstrate that significant improvements in performance can be made for both realistic channel coherence intervals and transmit SNR levels.

Appendix A
MMSE Channel Estimation Error

The minimum MSE matrix for the estimation of $\mathbf{h}$ can be written as

$$\begin{aligned}\varepsilon_{\mathrm{MMSE,est}} &= \mathbb{E}_{\mathbf{h},\mathbf{n}}\left[\left(\hat{\mathbf{h}}_{\mathrm{MMSE}}-\mathbf{h}\right)\left(\hat{\mathbf{h}}_{\mathrm{MMSE}}-\mathbf{h}\right)^H\right]\\ &= \sigma_H^2\left[\mathbf{A}_0^H\left(\mathbf{X}_T^H\mathbf{X}_T+\frac{\sigma_n^2}{\sigma_H^2}\mathbf{I}\right)\mathbf{A}_0-\left(\mathbf{A}_0^H\mathbf{X}_T^H+\mathbf{X}_T\mathbf{A}_0\right)+\mathbf{I}\right]\\ &= \sigma_H^2\left(\mathbf{I}-\mathbf{X}_T^H\left(\mathbf{X}_T^H\mathbf{X}_T+\frac{\sigma_n^2}{\sigma_H^2}\mathbf{I}_{n_T}\right)^{-1}\mathbf{X}_T\right)\\ &= \sigma_H^2\left(\mathbf{I}+\frac{\sigma_H^2}{\sigma_N^2}\mathbf{X}_T\mathbf{X}_T^H\right)^{-1}\\ &= \left(\sigma_H^{-2}+\frac{1}{\sigma_n^2}\frac{E_T}{M}\right)^{-1}\mathbf{I},\end{aligned} \tag{8}$$

where we have assumed that $\mathbf{h}$ and $\mathbf{z}$ are independent. The fourth equality follows from application of the matrix inversion lemma, $(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1}+\mathbf{DA}^{-1}\mathbf{B}\right)^{-1}\mathbf{DA}^{-1}$. Since the estimation error $\hat{\mathbf{h}}_{\mathrm{MMSE}}-\mathbf{h}$ is a linear combination of random vectors from a multivariate Gaussian distribution with uncorrelated components, it follows that the estimation errors are also independent Gaussian random variables.

Appendix B
Optimum Training and Data Energy Allocation

Here, we derive a closed-form expression for the optimum training energy E_T^* that minimizes the sum-MSE precoder design under a sum-energy constraint, $E_T + E_D \leq E_{\max}$. Due to space limitations, we are only able to show the most common case of long blocks (with $n \gg M$, and consequently $n_D > M$); however, the identical result applies for $n_D \leq M$.

We perform the optimization in terms of the training power $P_T = E_T/M$. Using the virtual uplink MSE from (3) as the objective function, and the energy constraints $E_T \geq 0$ and $E_T \leq E_{\max}$, we derive the Karush-Kuhn-Tucker (KKT) conditions

$$\frac{\partial \mathrm{SMSE}_{UL}}{\partial P_T}+\lambda_{\max}M-\lambda_+=0 \tag{9}$$

$$P_TM\geq 0, \qquad P_TM\leq E_{\max} \tag{10}$$

$$\lambda_+\geq 0, \qquad \lambda_{\max}\geq 0 \tag{11}$$

$$\lambda_+P_TM=0, \qquad \lambda_{\max}\left(P_TM-E_{\max}\right)=0. \tag{12}$$

We consider only the solutions where the constraints are not binding, as allowing either constraint to hold with equality prevents us from reaching a global minimum for the optimization problem. When $P_TM = 0$, no training symbols are sent, and the resulting channel estimate is $\hat{\mathbf{H}}^H = \mathbf{0}$. If $P_TM = E_{\max}$, zero energy remains for data transmission. In either of these cases, the resulting data symbol estimates are $\hat{\mathbf{x}}_k^{UL} = \mathbf{0}$, and no information can be communicated. Since neither constraint is binding, complementary slackness (12) requires that $\lambda_{\max} = \lambda_+ = 0$; thus, any minimizer can be found by considering the unconstrained minimization of SMSE_{UL} and checking feasibility of the resulting solutions. We begin by rewriting the effective noise power,

$$\sigma_{\mathrm{eff}}^2=\sigma_n^2+\frac{\sigma_n^2}{n_D}\left(\frac{E_{\max}-P_TM}{\rho+P_T}\right), \tag{13}$$

with $\rho = \sigma_n^2/\sigma_H^2$. Define the derivative

$$D_\sigma \doteq \frac{\partial\sigma_{\mathrm{eff}}^2}{\partial P_T}=\frac{-\sigma_n^2\left(E_{\max}+\rho M\right)}{\left(\rho+P_T\right)^2}. \tag{14}$$

We then separate the data power P_D from the uplink power allocation by rewriting $\mathbf{Q} = P_D\tilde{\mathbf{Q}}$, with associated sum power constraint $\mathrm{tr}\left[\tilde{\mathbf{Q}}\right] \leq 1$. It follows that

$$\tilde{\mathbf{R}}=\left(\frac{E_{\max}-P_TM}{n_D}\right)\mathbf{HV}\tilde{\mathbf{Q}}\mathbf{V}^H\mathbf{H}^H+\sigma_{\mathrm{eff}}^2. \tag{15}$$

Define the derivative of the trace function

$$\begin{aligned}D_{\mathrm{tr}} &\doteq \frac{\partial\mathrm{tr}\left[\tilde{\mathbf{R}}^{-1}\right]}{\partial P_T}=-\mathrm{tr}\left[\tilde{\mathbf{R}}^{-1}\frac{\partial\tilde{\mathbf{R}}}{\partial P_T}\tilde{\mathbf{R}}^{-1}\right]\\ &= \mathrm{tr}\left[\tilde{\mathbf{R}}^{-2}\left(\frac{M}{n_D}\mathbf{HV}\tilde{\mathbf{Q}}\mathbf{V}^H\mathbf{H}^H-D_\sigma\mathbf{I}\right)\right]\\ &= -\mathrm{tr}\left[\tilde{\mathbf{R}}^{-2}\right]\left(D_\sigma+\frac{M\sigma_{\mathrm{eff}}^2}{n_DP_D}\right)+\frac{M}{n_DP_D}\mathrm{tr}\left[\tilde{\mathbf{R}}^{-1}\right].\end{aligned} \tag{16}$$

The candidate values of P_T for unconstrained global optimality satisfy

$$\begin{aligned}\frac{\partial\mathrm{SMSE}_{UL}}{\partial P_T} &= D_\sigma\mathrm{tr}\left[\tilde{\mathbf{R}}^{-1}\right]+\sigma_{\mathrm{eff}}^2D_{\mathrm{tr}}=0\\ &= \left(\mathrm{tr}\left[\tilde{\mathbf{R}}^{-1}\right]-\sigma_{\mathrm{eff}}^2\mathrm{tr}\left[\tilde{\mathbf{R}}^{-2}\right]\right)\left(D_\sigma+\frac{M\sigma_{\mathrm{eff}}^2}{n_DP_D}\right).\end{aligned} \tag{17}$$

The first term in (17) can be rewritten as $P_D\mathrm{tr}\left[\tilde{\mathbf{R}}^{-1}\mathbf{HV}\tilde{\mathbf{Q}}\mathbf{V}^H\mathbf{H}^H\tilde{\mathbf{R}}^{-1}\right]$, which only has a trivial zero $P_T = E_{\max}/M$ (corresponding to $P_D = 0$) since the argument of the trace function is positive definite for non-zero power allocations $\mathbf{Q}$. Any globally optimum P_T^* must therefore satisfy

$$D_\sigma+\frac{M\sigma_{\mathrm{eff}}^2}{n_DP_D}=0. \tag{18}$$

Substituting the definitions of (13) and (14) gives rise to the following quadratic equation in P_T,

$$P_T^2\left(n_D-M\right)+2P_T\left(E_{\max}+\rho n_D\right)=\frac{E_{\max}^2}{M}-\rho^2n_D. \tag{19}$$

The two roots of this quadratic equation are

$$P_T=\frac{1}{n_D-M}\left(-E_{\max}-\rho n_D\pm\gamma\right), \tag{20}$$

with

$$\begin{aligned}\gamma &\doteq \sqrt{n_D\left(\rho^2 M + 2\rho E_{\max} + \frac{E_{\max}^2}{M}\right)} \\ &= E_{\max}\sqrt{\frac{n_D}{M}} + \rho\sqrt{n_D M}\end{aligned} \tag{21}$$

Clearly, for $n_D > M$, the negative root $(-\gamma)$ results in an infeasible solution $P_T < 0$. We can see that the positive root gives rise to

$$\begin{aligned}P_T^* &= \frac{E_{\max}\left(\sqrt{\frac{n_D}{M}} - 1\right) - \rho n_D\left(1 - \sqrt{\frac{M}{n_D}}\right)}{n_D - M} \\ &= \frac{E_{\max}\left(\frac{\sqrt{n_D}-\sqrt{M}}{\sqrt{M}}\right) - \rho n_D\left(\frac{\sqrt{n_D}-\sqrt{M}}{\sqrt{n_D}}\right)}{\left(\sqrt{n_D}-\sqrt{M}\right)\left(\sqrt{n_D}+\sqrt{M}\right)} \\ &= \frac{\frac{E_{\max}}{\sqrt{M}} - \rho\sqrt{n_D}}{\sqrt{n_D}+\sqrt{M}}.\end{aligned} \tag{22}$$

This solution always satisfies $P_T^* M < E_{\max}$, and is only infeasible (with $P_T^* < 0$) if $E_{\max} < \rho\sqrt{n_D M}$.

Finally, we prove that this stationary point P_T^* is indeed a global minimum. We observe that the second derivative of SMSE_{UL} can be written as

$$\begin{aligned}&\mathrm{tr}\left[\tilde{\mathbf{R}}^{-1}\mathbf{HVQV}^H\mathbf{H}^H\tilde{\mathbf{R}}^{-1}\right]\frac{\partial}{\partial P_T}\left(D_\sigma + \frac{M\sigma_{\text{eff}}^2}{n_D P_D}\right) \\ &+\left(D_\sigma + \frac{M\sigma_{\text{eff}}^2}{n_D P_D}\right)\frac{\partial}{\partial P_T}\left(\mathrm{tr}\left[\tilde{\mathbf{R}}^{-1}\mathbf{HVQV}^H\mathbf{H}^H\tilde{\mathbf{R}}^{-1}\right]\right),\end{aligned} \tag{23}$$

but the second term vanishes at P_T^* due to (18). We previously showed that the trace term is strictly positive; thus, to prove that P_T^* is a global minimizer, we must only show that the remaining term in the second derivative is positive at P_T^*:

$$\begin{aligned}\frac{\partial}{\partial P_T}\left(D_\sigma + \frac{M\sigma_{\text{eff}}^2}{n_D P_D}\right) &= \frac{\partial D_\sigma}{\partial P_T} + \frac{MD_\sigma}{n_D P_D} + \frac{M^2\sigma_{\text{eff}}^2}{n_D^2 P_D^2} \\ &= \frac{\partial D_\sigma}{\partial P_T} + \frac{M}{n_D P_D}\left(D_\sigma + \frac{M\sigma_{\text{eff}}^2}{n_D P_D}\right).\end{aligned} \tag{24}$$

At the point $P_T = P_T^*$, the second term vanishes due to (18). The remaining term

$$\left.\frac{\partial D_\sigma}{\partial P_T}\right|_{P_T = P_T^*} = \frac{2\sigma_n^2\left(E_{\max} + \rho M\right)}{\left(\rho + P_T^*\right)^3}, \tag{25}$$

is positive; thus, the training power P_T^* is the global minimizer. ■